THE ELECTROCARDIOGRAM IN CARDIOVASCULAR EPIDEMIOLOGY: PROBLEMS IN STANDARDIZED APPLICATION†

Henry Blackburn

Laboratory of Physiological Hygiene
University of Minnesota

The electrocardiogram (ECG) is employed in cardiovascular population studies because a) an objective record is obtained of a bio-electrical signal characteristic of the individual, b) it is of proven, if limited, value in diagnosis of certain specific and many non-specific anatomic and functional derangements of the heart muscle, c) its crude but significant predictive value is established for risk of future cardiac events and mortality, d) the information contained in the ECG is of a different nature from, and largely independent of that obtained by the medical history and physical examination, e) and the total yield of events relevant to the study of heart disease is increased by ECG recording.

These characteristics of the ECG suggest a particular suitability for application to cardiovascular epidemiology. In fact, however, conventional application of the ECG is in a quite primitive state. Problems basic to ECG theory as well as practical difficulties in instrumentation and method seriously limit its exploitation as an epidemiological study tool. Direct confrontation of these problems is required with attention both to broad concept and to fine detail.

INSTRUMENTATION

Direct paper-writing, amplifier-type electrocardiographs, internationally used, have differing fidelity of response, stability, speed and characteristics of writing which affect measurements and interpretation.1 Minimum specifications for frequency response2,3 are often not met by apparatus in use.4,5 These minimum standards are themselves not always adequate for faithful reproduction of the ECG in lower frequencies which contain known diagnostic information (i.e. the S-T segment and T wave).6,7 Frequencies exist above those obtained with these apparatus, but the amount and diagnostic significance of the information contained is uncertain.8

In multichannel apparatus which employ a common central terminal, little consideration has been made for accommodating high impedance sources at the skin-electrode contact, with attendant possibilities for noise, attenuation, and distortion.5,7 Other instrumentation problems exist, and provision is

†The material reported here was obtained with the support of grants from the U.S. Public Health Service (HE 03088, HE 04697, HE 04997), from the American Heart Association and the Minnesota Heart Association, and from the U.S. Public Health Service to the U. of Minnesota Medical School (HE 06314).
made in no type of recorder for simple built-in testing routines capable of operation by a technician.

In the dawning modern era of biomedical recording on magnetic tape, acquisition of good and faithful ECGs still presents many problems. Instruments and procedures in this field are inadequate and non-standard. Considerations preliminary to acquisition of ECG data on magnetic tape have been provided in outline by the Research Committee of the International Society of Cardiology.8+ 

The American Heart Association and the International Society of Cardiology have recognized the importance of good records both to clinical and epidemiologic practice and are now preparing recommendations intended to improve the quality of the machines and methods for their testing and to bring reason and standards into the field of instrumentation and procedure.

In the interim, a few constructive suggestions are possible concerning apparatus available for epidemiological use, though no instrument is satisfactory in all respects: 1) multichannel research-type direct-writer instruments contain higher grade components than the single channel machine produced for a mass market, and provide greater efficiency, fidelity, and stability, in recording. 2) Their frequency response, stability, behavior when overloaded, paper speed, and ruggedness in operation should be tested, independently of the supplier, as a routine prior to field use, with special attention given to frequency response at the lower end. 3) Collaborative studies seeking fullest opportunity for data comparison should employ the same model instrument. 4) Simultaneity of channels is required for orthogonal lead records. 5) The possibilities and difficulties of faithful recording of ECG's on magnetic tape

†(1) Competent engineering aid is required to set up and operate biomedical tape systems. (2) Control units for tape recorders must be designed in collaboration with those who will process the data. (3) Instrumentation in this field is still in development, no apparatus is entirely satisfactory, and many have serious limitations. (4) Specifications by suppliers, especially regarding noise level, are expressed variously and may be misinterpreted, while important characteristics may be omitted from descriptions. Satisfactory performance for the needs of the investigator, in testing after delivery, should be prerequisite to purchase. (5) All apparatus available give problems with noise, and methods for their solution involve sacrifices in data and may be costly. (6) All biomedical tape recorders available have satisfactory frequency response characteristics, at different costs, for electrocardiographic work, and most allow analysis of high frequency ECG components. (7) Time required for the machine to get up to operating speed must be considered in the light of the data processing method used and tape costs. (8) IRIG standard tape head configuration is a great practical advantage to collaborative collection-analysis programs. (9) Density of data recording at different sensitivities relates to the research needs and tape cost. (10) Preamplifiers designed for operation with specific tape units are unavailable at this time. (11) Input-output range of the recorder is important in regard to the type of data involved, as is behavior of the instrument when loads are imposed which exceed its operating range. (12) Input-output impedance, unity gain calibration, and an external calibration source are important concerns. (13) The number of channels, the ability to expand the system, and loss of data channels for noise compensation should be considered. (14) Magnetic tape of quality adequate for programs of automated computer recognition should be used.
and their subsequent classification should be explored by those commencing large population studies.

**TECHNIQUE**

Extraordinary demands on ECG technique are imposed by epidemiological application because of the necessity to record accurately and efficiently in large groups under restrictions of time and access to the subjects and to process masses of data. The most important characteristic of a technically good electrocardiogram, beyond that of fidelity, is freedom from baseline fluctuation and electrical "noise." Routine and newer procedures for reducing noise are provided in the appendix to this article and elsewhere but the most important single factor is preparation of the skin-electrode contact which need be standard only to the extent of being effective in reducing skin electrical impedance.

**Recording Conditions**

Some conditions of recording which affect the ECG are not susceptible of standardization in population studies. For example, meal effect on the S-T and T amplitude may be important and may be non-random, according to the total amount of coronary disease in a population. However it should be feasible in most surveys to minimize the acute effects of cold, smoking and heavy physical activity on heart rate and S-T and T wave amplitude by appropriate arrangement of the examination schedule and situation.

ECG recording supine is standard practice in dealing with sick patients, and ECG "normal standards" are based largely on supine records. This posture creates problems, to be discussed, in accurate siting of electrode positions and for the rationale in using orthogonal ECG leads and performing exercise ECG tests. However, complete information is required before there can be consideration of standard positions other than supine. Clearly, repetitions should be made in identical postures.

**The Record Itself**

At least three technically good complexes per lead, both at 25 and 50 mm. per second paper speed, are suggested for the resting ECG, and longer strips are obtained if the baseline is not stable or an arrhythmia is present. Duplicate records are desirable; a 1 mv. calibration should appear on each complete tracing, and serial number, study number, name and date are recorded on each. Measuring or reading time is reduced more than half by convenient mounting on stationery size cardboard (21.5 x 28 cm.), and pressure sensitive papers should be covered with cellophane. Records are best filed numerically and separately from other survey data.
Electrode Positions

All lead systems are sensitive to non-standard placement of chest electrodes, a source of inter-individual variation and of variation within subjects undergoing repeat procedures. The most common errors of placement in the "conventional" chest leads V$_1$ - V$_6$ are a) spacing between V$_1$ and V$_2$, usually too wide, b) misjudging the vertical references for V$_4$-V$_6$ often due to applying electrodes in the supine position, and c) placing V$_5$ and V$_6$ along the 5th intercostal space rather than in the horizontal plane of V$_4$. Special chest calipers, protractors, and templates are available to reduce variation in electrode placement.\textsuperscript{11,12,13}

Lead Systems

The widely used 12 lead "conventional" recording at rest (I, II, III, aVR, aVL, aVF, V$_1$, V$_2$, V$_3$, V$_4$, V$_5$, V$_6$) is now regarded as a minimum standard for epidemiological use.\textsuperscript{8,14} This fact presents as many problems as it solves. The convention is not accepted in a few countries, the number of leads is inconvenient, and the information contained in them is redundant and internally inconsistent. However, details concerning redundancy and wider evaluation of systems based on fewer leads are needed before the conventional one can be replaced. Improvement of the diagnostic power of the 12 lead arrangement obtains, even when conventionally recorded without benefit of phase relationships or high fidelity, if precise measurements are possible.\textsuperscript{15}

The potential of systems with 3 orthogonal leads based on theoretical models has been effectively demonstrated, including the possibility of reduced variability and sharpening of discriminative power.\textsuperscript{16} Their relevance to the problem of standard application in epidemiology is obvious, and adequate validation is awaited. The crucial pragmatic confrontation has not occurred between the results of accurate computer-assisted measurements and analysis of data from the same populations recorded both with conventional and orthogonal systems. Such studies are underway.

Standardization of orthogonal lead systems based on the equivalent dipole concept, is, for cardiovascular epidemiology, based on practical considerations. These systems, all imperfect, are not amenable to quantitative transformation into terms of each other, and their comparability in discrimination or classification in group studies is uncertain. One system is in widest use due primarily to its simple configuration of the smallest number of electrodes.\textsuperscript{17}

Measurement

Human measurement of even limited numbers of amplitudes and durations in scalar ECG tracings is quite impractical for epidemiological purposes. Mechanical attempts to digitize such records have not shown promise of accuracy
or efficiency. The only possibility for a standard base of precise, reproducible, efficacious measurement, and optimal use of detailed information in the continuous distributions of ECG values, lies in the computer methods described elsewhere in this symposium. When reliable measurements are obtained, adequate techniques are available to handle the numbers. ECG application to epidemiology would then need be concerned only with validation of the criteria for classification.

**PATTERN RECOGNITION**

Problems encountered in systematic study of human recognition and measurement of ECG waves are relevant both to current clinical ECG reading practice and to programming of computers for these purposes. Consider a few examples in definition, measurement and classification of diagnostic information in the ECG:

**Q and QS Waves**

The scalar and non-simultaneous time-based display of ECG potential changes was early accorded arbitrary labels in which the initial vectors in one lead were called a Q wave if the deflection was negative, whereas simultaneous forces in another lead, opposite in direction, were called an R wave. In the standard arrangements of leads I, II, and III, empirical observations led to an association of wide and deep Q waves in lead I with "anterior" infarction and in lead III with "posterior" infarction of the myocardium. Definition of Q, R, and S waves of the scalar display are the basis for conventional criteria of ECG diagnosis of myocardial infarction. The deflections, of course, are functions of the frequency response and paper speed of the recording system and their measurement is greatly affected by artefacts of the system. A sharp deflection is rendered a gradual one by the simple expedient of increasing the paper speed above the common 25 mm./sec., or by modifying the frequency response. FIGURE 1a indicates "ideal" measurement of a Q wave. Definition of these waves is arbitrary and empirical because the ECG baseline contains "noise" and is not a straight line. FIGURE 1b shows the logical but impractical alternative for measuring Q duration in writing systems in which the width of the trace baseline is important, here determined by the velocity of the deflection. The wide baseline may contain electrical "noise," but here it is wide for no other reason than custom (a wide hot stylus was early employed to mimic the wide baselines obtained with photographic records of the string galvanometer). The common practical manner of measuring Q duration in conventional records is seen in FIGURE 1c.

Minimum criteria are not established for the "simple" decision about presence or absence of a wave. Absence of a small initial R wave in chest leads V₂ to V₄ is associated with great probability of infarction of the heart. How high, how persistent, and at what slope must an initial deflection be to be
called an R wave? This sort of problem, illustrated in FIGURES 1d and e, is responsible for much observer variation. Beat-to-beat variation in the total presence or absence of a wave is unaccounted for in human or machine diagnostic criteria.

**S-T J Recognition**

The termination of QRS is another source of measurement error as the return to “resting state” is more gradual than depolarization and may actually
overlap with rising potential differences of repolarization of the heart muscle. The junction (J point) between QRS and the S-T (or R-T) segment lies along a gradual slope (FIGURE 1f). Since ischemic muscle repolarizes early, there is diagnostic information in this area of the trace but it is difficult to measure quantitatively. Examples of uncertainties in S-T segment and J point recognition are seen in FIGURES 1f, g, h, i, and j. FIGURE 1f is a common difficulty. Clearly all portions of the S-T are depressed below the P-R baseline, but the curvature of the segment nowhere allows consistent recognition of a junction by the human operator. FIGURE 1g appears to be a downward “sagging” S-T, but classification based on slope depends on where the J point is selected, whether at the sharp break near the base of the descending limb of R, or at a less distinct point. .08 sec. later. In FIGURE 1h the S-T-T has a distinctly “ischemic” appearance, but the two slopes, one positive, one negative, present problems in classification and in measurement of the appropriate negative displacement. FIGURE 1i illustrates at least four alternatives for location of the J point depending on definition of the end of QRS. FIGURE 1j is a particularly perplexing problem as this type of S-T segment is more often associated with “normal” than “abnormal” subjects, and correct measurement and classification require more precise definition of the end of QRS and of the contour of S-T.

Most of these problems may be reduced by clear procedural rules for the coder, but the detail required for successful standard application, for this and any clinical technique, is quite imposing.

CLASSIFICATION

Abundant evidence exists that agreement between clinical interpretations of the ECG for the frequency of “coronary heart disease,” or for presence or absence of ECG “abnormality,” is low, seriously limiting the clinical method for epidemiological application. The human interpretive process combines ECG forms and values in an unpredictable way so that comparisons, the essence of epidemiology, are frustrated. Recognition of these facts led to development of a system for classification of the conventional scalar ECG which aimed to state simply what is found in the record, concerning findings believed to be relevant to the epidemiology of heart disease. There is good general agreement about what items in the ECG to look for, but no previous attempt made to define them precisely.

THE MINNESOTA CODE

The attempt was made to codify departures of various types and degrees from ECG values determined from distributions in normal and abnormal subjects. The classes are based on clearly defined cut-off points which are the basis of commonly employed diagnostic criteria and provide a
reference frame applicable to the usual technique of the cardiographic reader. Actual coding of individual ECG items, requires, however, no diagnostic interpretation.

Other requisites of a useful classification scheme were partially met in the homogeneous groupings of items, a coding procedure giving mutually exclusive subclasses, and by extensive testing for reliability and yield of reasonable distributions in well-defined and in unknown populations.

The system has many weaknesses above those inherent in the redundancy and distortion and poor validation of conventional electrocardiography itself. The nature of such a clinical reading process necessitates loss of much information contained in detailed distributions of ECG values. Graphic illustrations were not provided of coding points and problem areas. Though the criteria are reasonable and generally accepted, coding rules and procedures are apparently inadequate for dealing with situations not clearly black or white, i.e. values around a cut-off level or tracings with beat-to-beat variation. As an individual may have several codable ECG findings, it is cumbersome to identify and tally numbers of persons; the axis of classification is around ECG items. Finally, an important degree of human observer variation persists even in such an "objective" attempt to characterize the "abnormal ECG," and even among trained coders.\textsuperscript{25 27}

\begin{table}
\centering
\caption{INTER-OBSERVER AGREEMENT ON Q WAVE CODING BY MINNESOTA CODE AMONG 1676 MEN AGES 40-60 BY TWO EXPERIENCED READERS IN ONE LABORATORY WITH EXCELLENT QUALITY RECORDS}

\begin{tabular}{|c|c|c|c|c|}
\hline
\textbf{Observer 1} & \textbf{I} & \textbf{2} & \textbf{3} & \textbf{0} & \textbf{T} \\
\hline
\textbf{I} & 13 & & & 13 & 13 \\
\hline
\textbf{2} & & 10 & 4 & & 14 \\
\hline
\textbf{3} & & & 7 & 3 & 10 \\
\hline
\textbf{0} & & & & 1636 & 1639 \\
\hline
\textbf{T} & 13 & 11 & 9 & 1643 & 1676 \\
\hline
\end{tabular}
\end{table}
Observer Variation in Minnesota Code Classification

TABLE 1 presents results for Q wave coding between two trained readers in the same laboratory utilizing records of good technical quality from a total population of 1676 middle-aged men. In numerous comparisons for individual ECG items under such circumstances, inter-observer agreement on the positive findings is approximately 80% for major Q, ST, T, block, and arrhythmia items, while intra-observer agreement for the same items and readers is around 90% (agreed presence of a finding/agreed presence + disagreed presence × 100).

TABLE 2
INTER-OBSERVER AGREEMENT ON Q WAVE CODING
BY MINNESOTA CODE AMONG 2158 MEN AND WOMEN
BY TWO TRAINED READERS IN SEPARATE LABORATORIES
WITH AVERAGE QUALITY RECORDS.‡

<table>
<thead>
<tr>
<th>Observer 1</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>0</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>15</td>
<td></td>
<td></td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>I_2</td>
<td></td>
<td>4</td>
<td>14</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>I_3</td>
<td></td>
<td></td>
<td>11</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>I_0</td>
<td></td>
<td>1</td>
<td>12</td>
<td>8</td>
<td>2079</td>
</tr>
<tr>
<td>I_T</td>
<td>20</td>
<td>26</td>
<td>20</td>
<td>2092</td>
<td>2158</td>
</tr>
</tbody>
</table>


A more representative situation is found in TABLE 2 in which two experienced coders in different laboratories read for Q waves in 2158 records of average quality from a large survey of adult men and women. Overall disagreement is doubled when three-way comparisons are made. The seriousness of observer variation is incontrovertible, though agreement on the major Q wave class is probably acceptable.
TABLE 3
INTER-OBSERVER VARIATION IN ECG CODING BY MINNESOTA CODE AMONG TRAINED PHYSICIAN, TRAINED CLERK, AND UNTRAINED CARDIOLOGIST GROUPS OF 3 OBSERVERS WITH ORDINARY QUALITY RECORDS.†

(Disagreement/agreement + disagreement x 100)

<table>
<thead>
<tr>
<th></th>
<th>Q waves</th>
<th>S-T depression</th>
<th>Negative T waves</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD Coders</td>
<td>74</td>
<td>47</td>
<td>25</td>
</tr>
<tr>
<td>Clerk Coders</td>
<td>60</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>Cardiologists</td>
<td>94</td>
<td>24</td>
<td>63</td>
</tr>
</tbody>
</table>


TABLE 3 is a study of intra-group variation among trained physician readers, trained non-medical clerks, and untrained readers who are expert cardiologists, in a test series of ECG photocopies of average technical quality sent out by W. H. O. A crude statistic is derived from the number of disagreements divided by the total of agreements and disagreements among the groups, according to the ECG item considered. Interesting findings of this detailed study are that training affects the consistency of coding practice and that trained female clerks with no knowledge of electrocardiography are more reliable readers in the sense of being reproducible than physicians who, it seems, know too much about the ECG to be reliable.

The phenomenon of like results, despite observer variation in classification of individuals, is illustrated in TABLE 4, in which independently assigned prevalence rates are given for individual ECG items in a large population, under the same conditions as those of TABLE 1. The generally similar frequencies obtained lead some to consider the classification system a useful epidemiological tool.

Application of the Minnesota Code continues, in the absence of better procedures, in many epidemiological study programs. Work is in progress, under auspices of W.H.O., the University Group Diabetes Program, the London School of Tropical Medicine and Hygiene, and others, to identify and correct for sources of observer variation in ECG coding with the system and to establish training and testing methods.

W.H.O. training and test records are now available from the Cardiovascular Section, W.H.O., Palais des Nations, Geneva along with detailed coding
TABLE 4
INDEPENDENTLY ASSIGNED PREVALENCE RATES (PER 1000 MEN) FOR ECG ITEMS IN A TOTAL POPULATION OF 1676 MEN, AGES 40-60 YEARS, BY TWO TRAINED READERS

<table>
<thead>
<tr>
<th>ECG Item</th>
<th>Minnesota Code</th>
<th>Observer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A (PR)</td>
</tr>
<tr>
<td>Large deep Q waves</td>
<td>1</td>
<td>7.2</td>
</tr>
<tr>
<td>Smaller Q waves</td>
<td>2,3</td>
<td>11.3</td>
</tr>
<tr>
<td>Left axis deviation $\geq 30^\circ$</td>
<td>1</td>
<td>40.6</td>
</tr>
<tr>
<td>High R waves (left)</td>
<td>3</td>
<td>128.9</td>
</tr>
<tr>
<td>S-T depression (rest)</td>
<td>1-4</td>
<td>38.8</td>
</tr>
<tr>
<td>Negative T waves</td>
<td>1-2</td>
<td>24.5</td>
</tr>
<tr>
<td>Prolonged P-R interval</td>
<td>3</td>
<td>22.1</td>
</tr>
<tr>
<td>Ventricular blocks</td>
<td>1,2,4</td>
<td>10.7</td>
</tr>
<tr>
<td>Significant arrhythmia</td>
<td>2-6</td>
<td>8.4</td>
</tr>
<tr>
<td>S-T depression (post-exercise)</td>
<td>1-4</td>
<td>40.6</td>
</tr>
</tbody>
</table>

procedures with illustrations, and suggestions for systematic continuous quality control among collaborating centers attempting to compare data. Absence of a calibrated standard for ECG and other clinical methods against which readers can be tested is unfortunate. A simple test-retest situation reveals only trends or systematic variations in classification. An attempt is in progress to provide an approximation to a standard with records of good quality in which reliable and precise wave measurements have been made by computer.

Application of the code has facilitated comparisons of rates of ECG abnormalities, and pointed up some areas worthy of special attention as a result of "unexpected" findings. "Revalidation" of the criteria employed has been, in a sense, obtained by establishing their individual predictive power for future cardiac events and mortality in the Framingham study material. Study of the inter-relations between ECG phenomena and other physical characteristics as seen in FIGURE 2 is rendered less cumbersome by such a classification scheme.
INTERPRETATION OF THE ECG IN EPIDEMIOLOGICAL APPLICATION

Questions about "standard values," validation of diagnostic criteria, and discrimination between "normal and abnormal" are among the more complex in biology. Electrocardiographic standards, such as those on which the Minnesota Code is based, have developed from empirical clinical-pathological observations and from distributions of ECG values found in "healthy" subjects. Most of the original observations derive from studies of less than ideal design, and, in fact, no ideal procedure is generally acceptable or even possible for establishing precise and valid standards.

Several considerations frustrate the attempt. The electrocardiogram may show no apparent (known) deviation in the presence of important myocardial involvement of several types. On the other hand, individuals with no manifest heart disease may exhibit identical ECG findings to those found in patients with clear clinical pictures of heart disease. Standards based on distributions of values even in well-defined polar groups give poor separation when applied to unknown populations due to the important degree of overlap between values for selected healthy subjects and cardiac patients.

There is a paucity of reference disease criteria independent of the electrocardiogram. A symptom such as chest pain has been used to "validate" ECG criteria but is at best a fuzzy reference as is heart size on chest x-ray, or the diagnostic impression of heart disease based on the history and physical examination. Even if there were the possibility of immediate direct study of the human heart after recording the ECG, no anatomic evidence would be
found for the majority of ECG anomalies which relate to myocardial function—including arrhythmias, conduction defects and negative S-T and T waves. Dead, injured, scarred and hypertrophied muscle, to a lesser extent coronary arterial occlusive lesions, are the only anatomic end points, and their classifications, in turn, are far from being reliable or quantitative. The distribution of ECG values found in well-described clinical heart disease classes is being established with large pathological studies using better, more standard methods, and there is some promise of useful independent references from cardiac catheterization and angiography in studies among living subjects.

Prospective epidemiological studies allow validation in both directions, i.e. assessing the predictive value of known ECG items for subsequent cardiac events and mortality, and also, looking backward after the event for clues in the characteristics of the earlier record. Newer computer-assisted techniques such as those reported in this meeting are finding diagnostic relationships in parts of the ECG complex not previously used in diagnosis. Auto- and cross-correlation studies are strengthening the power of electrocardiography.\(^{16,32,33}\)

Most reliance in ECG differentiation of “normal and abnormal” is now placed on ranges of values found in small, healthy, usually younger, groups. Some contribution to understanding of these problems with improvement in available standard values has been made by Simonson, with others from this laboratory.\(^{34,36}\) Despite questionable assumptions and theses, an approach was obtained to useful standard values by starting 1) with records from large groups of working people—large enough to be “reasonably representative” and to allow subsamples of sufficient size for statistical treatment—then 2) eliminating by methods independent of the ECG those physical conditions known to affect the ECG, and finally, 3) excluding rare “healthy” subjects with clear though “silent” patterns of ECG abnormality.

The ECG distributions obtained are often skewed.\(^{34}\) Skewness to the right is characteristic of many biological distribution curves. This does not necessarily mean that separate populations exist in the group but rather that there are more rigid biological restraints at one end of the distribution than at the other. However, the skewness in ECG distributions prevents effective use of measures of central tendency as standards. Simonson’s approach has been to employ the 2.5 and 97.5% range limits, rather than standard deviations, found in clinically “healthy” subjects, after brute force treatment of stray values. By definition, the extremes among the healthy subjects are abnormal. The greater probability of future disease among those at the extremes of the “normal” ECG distribution is borne out in follow-up study of minor non-specific Q, ST, and T findings.\(^{29,37}\)

But such standards applied for individual ECG items can result in untenable classifications in populations, since there is a strong chance that a healthy subject will have at least one ECG item, out of many dozens for which standards exist, exceeding the “normal” 95% limits. For example, in a recent national experiment, “healthy” subjects were excluded from participation.
on the basis of "abnormal" right or left axis deviation of a degree representing approximately the upper 2.5% limits in existing "standard values." Of course, it was found that approximately 5% of the apparently healthy men had to be eliminated at the outset since these two ECG items are completely independent. When similar upper limits are used for eight major classes of ECG finding, (including Q waves, axis, amplitudes, block patterns, S-T and T depression, conduction defects and arrhythmias) from 25 to 60% of total populations of middle-aged working men, depending on age and area, will be found to have "abnormal ECGs."

Finally a virtually untouched and complex problem in epidemiologic application of the ECG arises in attempts to evaluate changes for better or worse in heart patients on experimental secondary attack prevention programs. There is no guide to valid criteria for meaningful changes in an an already abnormal baseline record.

**THE EXERCISE ELECTROCARDIOGRAM**

Separate consideration is given to standardization of the exercise electrocardiogram because of the special problems and extraordinary confusion which exist in the field. The role of the exercise ECG in cardiovascular epidemiology is not established, as the method is inconvenient and time-consuming, and many questions remain about the methodology of the test and the significance of the stress response. Application has been made in group studies because of evidence that the post-exercise ECG response relates to subsequent disease risk and from interest in characterizing the circulatory fitness of populations. Experience about the safety of the various submaximal tests is now considerable.

Quite different methodologies have developed among and between those concerned with work physiology and those primarily interested in diagnostic use of the ECG response to work. In addition, recent claims, inadequately supported, that characteristics of the ECG recorded during a step test are of greater value than those after standard exercise have served to complicate further both methods and interpretation. Monitoring during ECG stress tests has been widely utilized for many years, and is highly recommended as a safety factor in earlier detection of gross arrhythmia or S-T segment depression while such tests are being applied in population studies.

Because of its frequency, the principal stress ECG response of known interest is behavior of the S-T segment and T wave. However, S-T and T amplitude and configuration, recognition, and interpretation are strongly dependent on 1) the fidelity of the recording instrument, 2) the ECG display in relation to baseline fluctuation and noise, 3) the lead configuration, 4) the type, amount, and duration of the work load imposed, 5) the period during and after the effort in which records are made, and on 6) the classification criteria or standards. The special problem lies in the fact that none of these variables is standardized.
FIGURE 3. Differences in magnitude and contour of S-T segment depression in one case recorded with different lead configurations (41). CH: forehead to chest, CR: right arm to chest, CC: right to left chest, CB: right back to chest, CM: sternum to chest, CS: right subclavicle to chest, O: right subclavicle to left back, X: right to left chest, A: manubrium to sacrum, B: right lower to left lower chest, CN: lower neck to chest, V: Wilson leads, RV: modified Wilson leads, R: right to left chest, L: ear-ensiform-precordial system.
The S-T and T diagnostic information is contained in the slower components of the ECG close to the range of the DC potential differences produced by motion at the electrode-skin contact. As previously mentioned, existing ECG apparatus may give unfaithful accords at these lower frequencies and most FM radio-telemetry devices on the market give stable records precisely because they are unresponsive to these slower components near the DC level. Non-linear "corrections" are employed in telemetry devices with short time-constants and distortions which may occur are unpredictable; the true curves cannot be re-synthesized. Solution to the problem of skin-electrode contact will allow in turn, improvement in instrument fidelity, and these are the first requisites to a standardized approach in this area. Meanwhile, cable-coupled recording on tested direct-writing apparatus is the safest policy for comparative studies.

Lead systems for exercise electrocardiography are almost as numerous as investigators in the field. In FIGURE 3 the simple self-evident demonstration is made that the display of diagnostic S-T segment information is dependent on the number of leads recorded and on location both of reference and exploring electrodes. Some systems appear more sensitive, perhaps more optimally distorted, for S-T display than others, and there are distinct differences between lead systems in regard to freedom from interference and motion artefact during work performance. Special electrodes and procedures are becoming available for reducing "noise" and for extracting the ECG signal from the "noise" of work records.

The most appropriate time intervals to record during and after work performance are not known and effective reduction of the vast amounts of data obtained from continuous recording is not yet attained. Solution of one problem relates to and awaits solution of another.

It would be desirable of course, to render similar both the aims and methods of those concerned with testing work performance and those interested in diagnostic comparisons and prognostic evaluation. There are differences in concept, however, which lead some investigators to an interest in the circulatory response to a given standard external work load, and others to concern with characterizing the ECG response to a standard physiological load imposed on the heart.

Bicycle work, or a step test "corrected" for body weight, defines the external work load but not the cardiac stress, does not allow standardized characterization of the ECG response of a population, and selectively "penalizes" the small individual. Constant rate step tests or treadmill work in a sense standardize for heart and body size but complicate the description of the physical work performed, and still yield considerable variability in energy cost between subjects, calculated as oxygen consumption per mass of heart muscle. A stress in the form of a particular submaximal procedure may represent little challenge to the heart in one population, while in another population a substantial number of men could not complete the same test.
A more rational basis for evaluating stress effects on the heart and circulation is provided by work against the individual’s own weight at aerobic work levels about his maximal $O_2$ consumption. However, application of such strenuous tests in large population surveys is not yet adequately tested and is subject to the same problems and variations as submaximal tests in regard to graded versus fixed work loads, continuous versus discontinuous procedure, different lead systems, postures, classification criteria, etc. It seems clear that such maximal tests increase the yield of ECG events of undoubted diagnostic and prognostic import over submaximal effort tests. Whether they significantly improve standardization for comparison of ECG response characteristics depends, among other things, on the selection involved (i.e. on the investigator’s judgment) in excluding subjects from the heavy test procedure and on the end points used for determining the maximum performance and for terminating the test procedure.

Broad discussions are now underway among several agencies to consider the current possibilities for standardization of exercise testing, and to point up the areas requiring study before more ideal standards are applicable. Reasonable and feasible first steps leading toward standardized application of the exercise electrocardiogram in descriptive and experimental cardiovascular epidemiology include:

1) A standard posture, preferably erect, throughout the baseline, work, and recovery periods.

2) A standard lead configuration throughout the baseline, work, and recovery periods, irrespective of the lead system chosen.

3) A graded procedure of submaximal work to include levels within the capacity of most subjects free of manifest heart disease.

4) Constant ECG monitoring with recording at intervals at most of one minute throughout and up to 4 minutes post-exercise.

5) Discussion leading to agreed criteria for exclusion from and termination of exercise tests, separately for diagnostic studies and for cardiac work evaluation studies.

6) A minimum criterion for a significantly positive ECG response of $\geq 0.1$ mv. horizontal or downward sloping S-T segment depression in any lead at any post-exercise interval recorded on an instrument with a minimum decay time constant of 2 seconds in a subject with no such finding in the baseline tracing.

Work required for a more rational basis for standardization in this area includes study of:

1) The relative ability of conventional ECG leads, orthogonal VCG leads, and highly distorted leads to display S-T depression and to discriminate between specified groups, i.e. “standards.”

2) The relative performance of the above leads during work procedures in regard to technical quality of the records.
3) The relative yield of events, discriminative and prognostic, employing single versus multiple work loads up to maximal tolerance.

4) Methods for extracting and reducing the amount of useful ECG information from inevitably noisy and prolonged work records.

5) Methods of analysis to improve the power of the test as an epidemiological tool.

6) Fundamental determinants of the ischemic electrocardiographic response, including the relationships to coronary flow, cardiac output, work, and oxygen consumption, and the role of physical conditioning.


Definitions and Baseline Reference

1)† The P wave is normally the gradual initial deflection of any group, and may be a summit or a depression. The level or reference from which its voltage is measured is the isoelectric level (T-P or U-P interval).

2)† The voltage of an upward QRS deflection is determined by measuring the vertical distance between the upper edge of the trace at the beginning of the QRS interval and the upper edge of the trace at the apex of the deflection. The voltage of a downward deflection is determined by measuring the vertical distance between the lower edge of the trace at the beginning of the QRS interval and the lower edge of the trace at the bottom of the deflection.

3)† The earliest QRS deflection which lies above the reference level should be labelled R. Any downward deflection which precedes R should be labelled Q. The first of any downward deflections which may follow R should be labelled S. The first of any upward deflections which may follow S should be labelled R', etc.

4)† When R is absent and the QRS complex consists of a single downward deflection, this deflection should be labelled QS. In statistical studies QS, Q, and S deflections should be considered separately.

5)† When the trace is descending it crosses the reference level at the instant when its lowest margin reaches a position below that which it occupied at the beginning of the QRS interval. When the trace is ascending, it crosses the reference level at the instant when its upper margin reaches a position above that which it occupied at the beginning of the QRS interval.

6)† When the form of the QRS complex varies from moment to moment because of the effect of respiratory movements on the position of the heart or for some similar reason, the classification of this complex should be determined by the variety of complex which is most abundant, or, if no type is numerically predominant, by the outline of the complexes which are of intermediate form.
7)† The term "S-T junction" or "J" (RS-T junction) should be used to indicate the point or shoulder which marks the end of the QRS complex, the point where the steep slopes of the QRS deflections are more or less abruptly replaced by the more gradual slopes which precede or comprise the first limb of the T wave.

8)† The reference level for the measurement of the displacement of the S-T junction (or J) should be the P-R segment at the beginning of QRS. The level of reference for the measurement of the S-T segment, the T wave and the U wave should be the isoelectric level (T-P or U-P interval) when this can be determined; otherwise it should be the level of the trace at the beginning of the QRS interval.

9)† The term "diphasic T waves" should be applied to those final ventricular deflections which present two distinct turning points, one on each side of the level of reference.

10)† The P-R interval is measured from the beginning of the P wave to the beginning of QRS whether this be represented by a Q wave or an R wave. It is recommended that the longest P-R interval found in the bipolar or unipolar extremity leads be regarded as the P-R interval.

11)† The QRS interval is measured from the beginning (Q or R) to the end of the QRS group of deflections. The longest QRS interval found in the bipolar or unipolar extremity leads is regarded as most nearly correct.

12) When in doubt, code to the less abnormal category. If there is a reasonable possibility that an abnormality may be artefactual, do not code it. Presence or absence of an item should be determined by the finding in the majority of recorded beats.

13) QS versus RS. At 25 mm./sec. paper speed with conventional apparatus an initial spike or sharp deflection of .025 mv. (1/4 mm.) or more is taken as an R. Examples of an acceptable R, and an unacceptable R, are given in FIGURES 1d and e in the test.

14) For coding, a Q wave must be 0.1 mv. in amplitude associated with an R wave of 0.1 mv.

15) S-T depression is placed in a questionable category or not tabulated when found in less than majority of beats in a lead, when the baseline is swinging wildly or sloping up or down, and when the S-T segment describes an arc with no clear junction with QRS. If the baseline is stable over three complexes take these in precedence over others. Post-exercise S-T depression is taken as the absolute value, not as a change relative to the resting level.

If more than one change of slope makes S-T, J point determination arbitrary, choose the latest one. Measure S-T, J and segment depression from the level of a horizontal tangent to the upper edge of baseline at the onset of QRS to the upper edge of S-T. S-T cannot be considered horizontal or downward sloping if any part of it slopes upward.

16) T wave is coded as positive if it is not definitely flat or negative. Negative T amplitude is measured from the lower edge of baseline at T-TP
junction. Where S-T-T demarcation is not clear, measure the amplitude of T inversion to the lowest level of the S-T-T segment. The positive phase of negative-positive diphasic T waves is ignored in coding.

17) R peak duration is measured from the onset of QRS (whether a Q or R) to the peak of the R wave.

18) For heart rate, average the R-R intervals from leads I and V₆, or the first and last resting leads recorded.

19) If in the majority of beats in any bipolar or unipolar limb lead P-R is 0.22 seconds or over, code prolonged P-R.

20) If in the majority of beats in any bipolar or unipolar limb lead QRS is 0.12 seconds or more, a ventricular conduction defect is coded.


Precordial Lead Placement

† Subscript 1 (V₁) shall be used for a lead from the right sternal margin at the fourth intercostal space; subscript 2 for a lead from the left sternal margin at the fourth intercostal space; subscript 4 for a lead from the fifth intercostal space where it is crossed by the midclavicular line; subscript 3 for a lead from a point midway between points 2 and 4; subscript 5 for a lead from the junction of the left anterior axillary line with the horizontal position of position 4; subscripts 6, 7, and 8 are for leads on the same horizontal level but at the left midaxillary line (6), the left posterior axillary line (7) and the left midscapular line (8), respectively.


Capacitive Interference

1) Electrode - Skin Preparation:
   Careful preparation of all skin-electrode contacts, rubbing to skin hyperemia, cleaning and drying with alcohol-ether or acetone.
   Application of “good quality” electrolyte jelly or paste (based on resistance measurements if newer creme type pastes used).
   Good, stable mechanical contacts.
   Electrodes cleaned with sandpaper and solvents, never steel wool.
   (High impedance cleaned at one electrode-skin contact can cause A-C pickup and baseline fluctuations.)

2) Lead Cables:
   Daily check of cables with an Ohmmeter. (Intermittent problems often due to loose or broken contacts in lead cables.)
3) Preamplifiers:
   Use differential preamplifiers for common mode rejection. (Transistorized pre-amplifiers tend to be unstable if D.C. drifts or source resistance are great.)

4) Grounding:
   Concentric grounding, avoiding ground loops. Use separate ground cable from subject to system ground.
   Ground system to good earthy ground.
   Check that zero line of AC power line is connected.

5) Shielding:
   Recording room containing or adjacent to as few power lines as possible, symmetric location of subject with respect to sources of interference (equidistant to electrodes).
   When necessary use shielding blanket (inside insulator, outside good conductor) under mattress and over subject. Ground conducting layer to system ground.
   Faraday cage is optimal protection from AC pick-up.

**INDUCTIVE INTERFERENCE**

Avoid cable coiling at input side, and place patient cables in straight lines along the subject.
Use shielded cables.
Careful selection of recording area, as far from disturbance as possible (i.e. magnetic fields from transformers, etc.).

**INTERFERENCE FROM MUSCLE ACTION, TREMORS, AND RESPIRATION**

Conventional techniques are applied.
For records during exercise, selection is made of contact sites relatively free of muscle movement and contact with moving extremities, clothing, etc. (Quiet areas are the sternum, and immediately medial to deltoid below clavicles. Intercostal spaces may be superior to over-the-rib position.)

**FILTERING**

Use of low pass filters in commercial electrocardiographs is inadvisable as they cause distortion of high frequency components and attenuated QRS amplitude. If used, record a few beats unfiltered.

**AVERAGING TO “SMOOTH OUT” NOISE**

Several techniques are available including small computers for average transients. Use should not replace efforts for best recording technique.
NOISE PECULIAR TO MAGNETIC TAPE RECORDINGS

Careful check-out that recorders meet all specifications. Basic noise level of many commercial systems is far from satisfactory.
Quality tape is required.
Temperature and humidity control and dust-free conditions are sought.
Clean recording heads frequently.

Bibliography


